

CONTRAT DE COLLABORATION ET DE RECHERCHE

Le **CENTRE NATIONAL de la RECHERCHE SCIENTIFIQUE**, Etablissement Public à caractère Scientifique et Technologique, dont le siège social est 3, rue Michel Ange, 75794 PARIS Cedex 16, n° SIREN 180 089 013, n° TVA Intracommunautaire : FR 40 180 089 013, code APE 7219Z, représenté par son Président-Directeur Général, Monsieur Antoine PETIT, lequel a délégué sa signature pour la présente convention à Monsieur Ludovic HAMON, Délégué Régional pour la Circonscription Centre Limousin Poitou Charentes, 3E avenue de la Recherche Scientifique, CS 10065, 45071 ORLEANS Cedex 2.

Le **CNRS** au nom et pour le compte de l'**UNIVERSITE DE POITIERS**, Etablissement Public à caractère Scientifique, Culturel et Professionnel, dont le siège social est situé 15, rue de l'Hôtel-Dieu, TSA 71117, 86073 POITIERS Cedex 09, France, n° SIRET : 19860856400375, code APE 8542Z, représentée par sa Présidente, Madame Virginie LAVAL,
Ci-après désignée par « l'UP »

Le CNRS et l'Université de Poitiers agissant pour son compte et le compte du Laboratoire de Mathématiques et Applications (LMA UMR7348), dirigé par M Boris PASQUIER, ci-après le « LABORATOIRE »

Et

ONAOS, Société par Actions Simplifiée, dont le siège social est situé au Cap Oméga, Rpt Benjamin Franklin, 34960 Montpellier, France, Siret : 83984908000024, code APE 6201Z, représentée par son Président, Monsieur Guilhem MAYORAL,
Ci-après dénommée la « **SOCIETE** »

Le CNRS, l'UP sont désigné ci-après les **ETABLISSEMENTS**.

Les **ETABLISSEMENTS** et la **SOCIETE** sont, ci-après, conjointement désignés par « **PARTIES** » et individuellement par « **PARTIE** ».

Les termes employés au singulier s'entendent aussi au pluriel et vice versa, selon le contexte.

PREAMBULE

Dans le cadre d'une convention CIFRE n° 2021/0861 (Convention Industrielle de Formation par la Recherche), la **SOCIETE** a embauché M Nassar DAKKOUNE, ci-après désigné par le « Doctorant-CIFRE ».

Pour ce faire, la **SOCIETE** souhaite organiser une collaboration de recherche avec les **ETABLISSEMENTS**.

La **SOCIETE** demande aux **ETABLISSEMENTS**, qui acceptent, d'accueillir le Doctorant-CIFRE au **LABORATOIRE** et de mettre à sa disposition les moyens scientifiques nécessaires à la préparation de sa thèse dans les conditions du présent contrat.

ARTICLE 1 – DEFINITIONS

ADAPTATION : tout logiciel résultant des modifications, corrections, traductions, adaptations ou évolutions mineures intégrées dans le code source d'un logiciel **RESULTAT** ou d'une **CONNAISSANCE PROPRE**, intégrant

de nouvelles fonctionnalités non séparables d'un tel logiciel. Les ADAPTATIONS appartiennent au propriétaire du logiciel à partir duquel elles sont réalisées.

AFFILIEES : le terme AFFILIEES désigne les entités, qui jouissent de la personnalité morale ou non, qui contrôlent ou sont contrôlées par une PARTIE ou sont sous contrôle conjoint des PARTIES. Aux fins de la présente définition, on entend par "contrôle" la détention directe ou indirecte d'au moins cinquante pour cent (50 %) des droits de vote, du capital ou d'autres titres d'une entité contrôlée ou contrôlée conjointement ou si elle possède, directement ou indirectement, le pouvoir d'orienter ou déterminer l'orientation de la direction et des politiques de l'entité, ou le pouvoir de choisir ou de désigner plus de cinquante pour cent (50 %) des membres de l'organisme de direction de l'entité ;

Pour le CNRS, AFFILIÉE signifie la société CNRS Innovation.

CONNAISSANCES PROPRES : toutes les informations et connaissances techniques et/ou scientifiques et/ou tout autre type d'informations, notamment le savoir-faire, les secrets de fabrication, les secrets commerciaux, les prototypes, les données, les bases de données, logiciels, ADAPTATIONS, les dossiers, les plans, les schémas, les dessins, les formules, sous quelque forme qu'elles soient, brevetables ou non et/ou brevetées ou non, ainsi que tous les droits y afférents, appartenant à une PARTIE ou détenue par elle avant la date d'entrée en vigueur du CONTRAT et/ou développées ou acquises par elle indépendamment de l'exécution de celui-ci.

Les CONNAISSANCES PROPRES de chaque PARTIE restent leurs propriétés respectives.

Les CONNAISSANCES PROPRES des PARTIES, nécessaires à l'exécution du CONTRAT sont précisées à l'Annexe 1 qui sera mise à jour au fur et à mesure.

DOMAINE D'EXPLOITATION : le DOMAINE D'EXPLOITATION est : Secteur santé/médical/paramédical, Application avec intelligence artificielle.

ETUDE : objet des recherches du CONTRAT détaillées en Annexe 2.

EXTENSION : tout logiciel permettant de réaliser des fonctionnalités ou services supplémentaires à ceux fournis par le logiciel à partir duquel l'extension est réalisée, de telle sorte que ledit logiciel reste exécutable de manière indépendante et que l'extension s'exécute en faisant appel au logiciel premier, l'un appelant l'autre au moment de leur exécution.

LOGICIEL NOUVEAU : logiciel créé *ex nihilo* dans le cadre du CONTRAT.

MANDATAIRE : PARTIE désignée par les ETABLISSEMENTS en application de l'article L.533-1 du code de la recherche.
Le MANDATAIRE est le CNRS.

RESULTATS : toutes les informations et connaissances techniques et/ou scientifiques et/ou tout autre type d'informations, notamment le savoir-faire, les secrets de fabrication, les secrets commerciaux, les prototypes, les données, les bases de données, les LOGICIELS NOUVEAUX et EXTENSIONS, les dossiers, les plans, les schémas, les dessins, les formules, sous quelque forme qu'elles soient, brevetables ou non et/ou brevetées ou non, ainsi que tous les droits y afférents, développées par une ou plusieurs PARTIES dans le cadre du CONTRAT.

ARTICLE 2 - OBJET

Le CONTRAT a pour objet de définir les modalités de la collaboration entre les PARTIES relatives à la réalisation de l'ETUDE.

Les PARTIES mettront tout en œuvre pour assurer son bon déroulement conformément à l'obligation de moyens qui leur incombe.

La SOCIETE emploie le Doctorant-CIFRE et s'engage à lui fournir toutes les facilités pour préparer une thèse sur le sujet défini au présent article, dans ses locaux.

Le travail de recherche confié au Doctorant-CIFRE défini d'un commun accord entre la SOCIETE, les ETABLISSEMENTS, et le directeur de thèse, portera sur l'ETUDE.

ARTICLE 3 – SUIVI DE L'ETUDE

Les PARTIES s'accordent sur un planning de réunions scientifiques de suivi en début de CONTRAT. Ce planning sert de guide et pourra être adapté au fur et à mesure de la durée de collaboration. Ces réunions pourront avoir lieu chaque fois qu'une des PARTIES l'estimera nécessaire.

Le suivi de l'étude sera assuré par les Responsables Scientifiques de l'ETUDE, Yousri Slaoui du LMA. Leur correspondante dans la SOCIETE est Mme Namrata PATEL.

Lorsqu'il se trouve au sein du LABORATOIRE, le Doctorant CIFRE est placé sous l'autorité administrative de la Directrice du LABORATOIRE et doit se conformer au règlement intérieur du LABORATOIRE. La SOCIETE continue toutefois d'assurer à son égard toutes les obligations civiles, sociales et fiscales incombant à l'employeur.

Pour la bonne exécution du travail de recherche qui lui est confié, le Doctorant-CIFRE passera 60 % de son temps de travail dans le laboratoire de la SOCIETE et 40 % de son temps de travail au LABORATOIRE (notamment pour les réunions mensuelles de suivi de l'avancement de thèse).

Les réunions font l'objet de comptes rendus qui sont transmis à chacune des PARTIES dans les quinze (15) jours suivant la date de la réunion.

Ce compte rendu est considéré comme accepté par les PARTIES si, dans les huit (8) jours à compter de la réception de ce compte rendu, aucune objection, ni revendication, n'a été formulée par écrit.

Par ailleurs le Responsable scientifique de l'ETUDE du LABORATOIRE et son correspondant au sein de la SOCIETE établissent des rapports intermédiaires de l'étude aux échéances suivantes :

- à T0 +12
- à T0 +24,

et un rapport final de synthèse dans le mois qui précède l'expiration ou la résiliation anticipée du CONTRAT.

ARTICLE 4 – MODALITES DE FINANCEMENT

La SOCIETE rembourse, les frais de missions du Responsable scientifique du LABORATOIRE et de ses collaborateurs affectés à l'ETUDE, qui auront été décidés d'un commun accord entre les PARTIES.

En contribution à l'ETUDE, la SOCIETE s'engage à verser au CNRS une somme détaillée en Annexe 3 d'un montant global et forfaitaire de :

- Montant HT : 30 000 Euros
- TVA 20 % : 6 000 Euros
- Montant TTC : 36 000 Euros

En cas de modification du taux de la TVA, il sera appliqué le taux en vigueur à la date de la facturation.

Cette somme est versée par la SOCIETE au compte n° 00001000035, Code banque 10071, Code guichet 45000, Clé RIB 89, ouvert au nom de l'Agent Comptable Secondaire de la Délégation Centre Limousin Poitou-Charentes du CNRS, avec la mention [n° 247473-ONAOS-LMA], sur présentation de factures aux échéances suivantes :

- 10 000 Euros HT à la signature du CONTRAT
- 10 000 Euros HT, douze (12) mois après la signature du présent contrat,
- 10 000 Euros HT, vingt-quatre (24) mois après la signature du présent contrat

Les factures sont adressées à la SOCIETE à l'attention de Mme. Namrata PATEL ou M. Guilhem MAYORAL.

La contribution HT de la SOCIETE représente donc 26 % du montant total HT de l'ETUDE estimée, conformément à l'Annexe 3, à 113 000 Euros HT.

L'emploi par le CNRS de la contribution forfaitaire versée par la SOCIETE n'est pas subordonné à des conditions de délai, ni à fourniture de justificatifs.

De convention expresse, les PARTIES écartent toute faculté pour la SOCIETE de procéder à la réduction de la somme détaillée en Annexe 3 au motif que l'exécution du CONTRAT ne lui donnerait pas complète satisfaction, quelle qu'en soit la cause.

ARTICLE 5 - PROPRIETE ET ACCES AUX RESULTATS

Il est convenu entre les Parties que les RESULTATS appartiennent aux PARTIES à parts égales, sauf le cas où les frais engagés par la Société dépassent 50% du cout global de l'Etude, auquel cas la répartition de la propriété se fera en tenant compte des apports financiers, humains et intellectuels de chaque Partie.

Le MANDATAIRE et la SOCIETE concluront un règlement de copropriété préalablement à l'exploitation de tout RESULTAT.

La SOCIETE prend en charge les frais de protection des RESULTATS.

ARTICLE 6 - UTILISATIONS AUX FIN DE RECHERCHE

Chaque PARTIE peut utiliser librement et gratuitement les RESULTATS dans le cadre d'actes accomplis à des fins de recherche à l'exclusion de toute activité à caractère industriel et/ou commerciales, à des fins de recherche..

ARTICLE 7 - EXPLOITATION DES RESULTATS

7.1 Dans le DOMAINE D'EXPLOITATION

Dans le DOMAINE D'EXPLOITATION et sous les réserves définies au présent article, la SOCIETE jouit d'un droit d'exploitation exclusif des RESULTATS.

Les ETABLISSEMENTS n'exploiteront pas les RESULTATS dans le DOMAINE D'EXPLOITATION. De ce fait, la SOCIETE s'oblige à exploiter à des fins commerciales, directement ou indirectement, les RESULTATS dans le DOMAINE D'EXPLOITATION et à verser une rémunération dont la nature et le mode de calcul seront définis d'un commun accord ultérieurement qui tiendra compte des frais engagées pour l'obtention des résultats, de la maturité des Résultats ainsi que pour le dépôt et le maintien des titres de propriété intellectuelle issue des Résultats.

7.2 Hors du DOMAINE D'EXPLOITATION

Hors du DOMAINE D'EXPLOITATION, les ETABLISSEMENTS ont l'exclusivité des droits d'exploitation des RESULTATS, et peuvent négocier librement tout contrat de licence d'exploitation portant sur ces RESULTATS.

Les ETABLISSEMENTS versent à la SOCIETE une part des redevances perçues, le cas échéant, au titre de l'exploitation dont la nature et le mode de calcul seront définis ultérieurement.

7.3 CONNAISSANCES PROPRES nécessaires à l'utilisation des RESULTATS

Une PARTIE ne reçoit aucun droit sur les CONNAISSANCES PROPRES de l'autre PARTIE du fait du CONTRAT.

Si l'exploitation des RESULTATS par une PARTIE nécessite l'utilisation des CONNAISSANCES PROPRES détenues pour partie ou en totalité par une autre PARTIE, celle-ci s'efforce de favoriser cette exploitation, sous réserve des droits consentis à des tiers au jour de la signature du CONTRAT ou qui pourraient être consentis pendant la durée du CONTRAT. Les conditions d'utilisation des CONNAISSANCES PROPRES sont alors fixées contractuellement au cas par cas.

ARTICLE 8 – CONFIDENTIALITE

8.1 Informations confidentielles

Une PARTIE (ci-après la « PARTIE DIVULGATRICE ») pourrait être amenée à divulguer à une autre PARTIE (ci-après la « PARTIE RECEPTRICE ») des informations confidentielles.

Les PARTIES conviennent que sont confidentielles toutes les informations divulguées par l'une des PARTIES à l'autre incluant notamment les CONNAISSANCES PROPRES (ci-après dénommées « INFORMATIONS CONFIDENTIELLES »), quel qu'en soit l'objet, la nature, le support et le mode de transmission, sous réserve que :

- si elles sont transmises sur un support, elles soient désignées comme « information(s) confidentielle(s) » de la PARTIE DIVULGATRICE par l'apposition ou l'adjonction sur leur support d'un tampon « Confidentiel » ou de toute autre mention appropriée, compréhensible par les PARTIES et adaptée au support ;
- si elles sont transmises oralement, le caractère d' « information(s) confidentielle(s) » ait été porté à la connaissance de la PARTIE RECEPTRICE au moment de leur divulgation et consigné comme tel dans le compte rendu de réunion au cours de laquelle l'information a été divulguée, ou en cas d'impossibilité, confirmé par écrit dans les trente (30) jours de la divulgation, étant entendu qu'entre-temps ces informations devront être traitées par la PARTIE RECEPTRICE comme des INFORMATIONS CONFIDENTIELLES.

Les PARTIES reconnaissent que les RESULTATS ne sont pas considérées comme des INFORMATIONS CONFIDENTIELLES au sens du présent article. En revanche les modalités de leur divulgation sont régies par l'article « Publications ».

Chaque PARTIE RECEPTRICE s'engage, pendant la durée du CONTRAT et pour une période de cinq (5) ans à compter du terme contractuel prévu ou de la résiliation du CONTRAT, à ce que les INFORMATIONS CONFIDENTIELLES qu'elle reçoit :

- a) soient protégées et gardées confidentielles et soient traitées avec le même degré de précaution et de protection que la PARTIE RECEPTRICE accorde à ses propres INFORMATIONS CONFIDENTIELLES de même importance ;
- b) ne soient divulguées de manière interne qu'aux seuls membres de son personnel ayant à en connaître, dûment informés du caractère strictement confidentiel de ces INFORMATIONS CONFIDENTIELLES, et ne soient utilisées par ces derniers que dans les conditions définies par le CONTRAT. Chacune des PARTIES déclare avoir pris ou s'engage à prendre les mesures nécessaires auprès de son personnel pour lui permettre de respecter les engagements pris au titre du CONTRAT ;
- c) ne soient pas utilisées, totalement ou partiellement, pour tout autre but que l'ETUDE, sans le consentement préalable et écrit de la PARTIE DIVULGATRICE ;
- d) ne soient divulguées, ni susceptibles d'être divulguées, soit directement, soit indirectement, à tous tiers, notamment aux sous-traitants ou à toutes autres personnes, sans l'autorisation préalable et écrite de la PARTIE DIVULGATRICE et, en cas d'autorisation de la PARTIE DIVULGATRICE, à la condition que le tiers bénéficiaire s'engage au préalable et par écrit à se soumettre aux mêmes obligations de confidentialité que celles contenues dans le CONTRAT ;

e) ne soient ni copiées, ni reproduites, ni dupliquées totalement ou partiellement sans l'autorisation préalable et écrite de la PARTIE DIVULGATRICE.

f) ne soient pas utilisées de manière à obtenir un quelconque droit de propriété intellectuelle (notamment brevet, marque...) par la PARTIE RECEPTRICE ou tout autre tiers dans quelque pays que ce soit.

8.2 Obligation de divulgation

La PARTIE RECEPTRICE s'engage à informer, par écrit et sans délai, la PARTIE DIVULGATRICE, si elle fait l'objet, en vertu de la loi ou dans le cadre d'une procédure judiciaire ou administrative, d'une demande de divulgation de toute INFORMATION CONFIDENTIELLE de la PARTIE DIVULGATRICE. Cette notification ne peut être interprétée comme une autorisation de la part de la PARTIE DIVULGATRICE à divulguer ces INFORMATIONS CONFIDENTIELLES.

8.3 Science ouverte

Dès que cela sera possible eu égard aux dispositions relatives aux INFORMATIONS CONFIDENTIELLES et à la protection et l'exploitation des RESULTATS et sous réserve du respect des modalités de communication prévues à l'article 9 « Publications », les PARTIES s'efforceront de diffuser largement au public l'information scientifique issue de l'ETUDE.

Le chef de projet de chaque action tiendra un plan de gestion des données afin de définir ce qui devra rester confidentiel et pour quelle durée (voir article « Publications »), les conditions d'archivage des données et informations relatives au projet et les informations et données qui pourront être diffusées au public ainsi que les modalités de cette diffusion.

Cette diffusion ne devra en tout état de cause pas faire obstacle à la protection des RESULTATS par un droit de propriété intellectuelle et, le cas échéant, par la délivrance d'un titre de propriété industrielle.

ARTICLE 9 - PUBLICATIONS

9.1 Communication

Toute publication ou communication des RESULTATS par l'une des PARTIES, devra recevoir, pendant la durée du CONTRAT et les 12 mois qui suivent son expiration, l'accord écrit des autres PARTIES qui feront connaître leur décision dans un délai maximum de 3 mois à compter de la demande. Passé ce délai et faute de réponse, l'accord sera réputé refusé.

Tout projet de publication sera soumis à l'avis des autres PARTIES qui pourront proposer des modifications sous réserve que cela soit justifié au regard de l'exploitation industrielle et commerciale des RESULTATS. Toutefois, si des modifications ont lieu, celles-ci ne doivent pas porter atteinte à la valeur scientifique de la publication.

Si des informations contenues dans la publication ou communication doivent faire l'objet d'une protection au titre de la propriété industrielle, une des PARTIES pourra retarder la publication ou la communication d'une période maximale de 18 mois à compter de la demande.

Ces publications et communications devront mentionner le concours apporté par chacune des PARTIES à la réalisation de l'ETUDE.

Ces stipulations ne pourront faire obstacle :

- ni à l'obligation qui incombe à chacune des personnes participant à l'ETUDE de produire un rapport d'activité à l'organisme dont elle relève, dans la mesure où cette communication ne constitue pas une divulgation au sens des lois sur la propriété industrielle ;
- ni à la soutenance de diplôme des chercheurs et ingénieurs, cette soutenance devant être organisée chaque fois que nécessaire de façon à garantir, tout en respectant la réglementation universitaire en vigueur, la confidentialité des INFORMATIONS CONFIDENTIELLES et des RESULTATS objets d'un Dossier Technique secret (DTS), tels que visés en article 8 et 9.1.

9.2 Dossier technique secret (DTS).

A titre exceptionnel, les RESULTATS pour lesquels l'une des PARTIES estime devoir appliquer une protection par secret donneront lieu à la constitution d'un dossier technique secret.

Dans ce cas, chaque PARTIE déterminera :

- la part des RESULTATS qui constituera ledit DTS, et qui par conséquent, ne pourra ni être publiée ni divulguée à un tiers sans son autorisation ;
- le contenu des informations qui ne relèveront pas du DTS et qui pourront faire l'objet d'une publication ou d'une communication à un tiers.

Les PARTIES détermineront d'un commun accord, la durée pendant laquelle le DTS restera secret et le retour financier devant être versé aux PARTIES devant s'abstenir de communiquer ou divulguer les RESULTATS.

ARTICLE 10 – DONNEES PERSONNELLES

Les PARTIES veillent au respect du Règlement européen de protection des données n°2016/679 du 27 avril 2016 (RGPD), de la loi n°78-17 modifiée du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (loi Informatique et libertés) et de toute réglementation nationale prise en application, concernant les traitements de données à caractère personnel mis en œuvre dans le cadre du présent CONTRAT.

Lorsque la réalisation de l'ETUDE suppose le traitement de données à caractère personnel autres que les données des personnes en charge de l'exécution contractuelle, une annexe définissant les engagements des PARTIES, les modalités techniques de mise en œuvre du traitement et la responsabilité, sera jointe au présent CONTRAT.

ARTICLE 11 – AFFILIEES

11.1 Confidentialité

Les PARTIES conviennent que la communication d'INFORMATIONS CONFIDENTIELLES ou de RESULTATS aux AFFILIEES est libre, sous réserve du respect par ces dernières de l'obligation de confidentialité prévue à l'article « Confidentialité ».

Chaque PARTIE se porte garante du respect de l'obligation de confidentialité prévue à l'article 8 par ses AFFILIEES.

11.2 Utilisation à des fins de recherche, exploitation

Le droit d'utiliser des RESULTATS ou des CONNAISSANCES PROPRES à des fins de recherche ou de les exploiter pourra être accordé aux AFFILIEES :

- à des conditions justes et raisonnables ;
- sur la base d'une entente bilatérale écrite ;

- sous réserve des droits consentis à des tiers au jour de la demande de l'AFFILIEE.

ARTICLE 12 – USAGE NOMS ET MARQUES

Chaque PARTIE autorise les autres PARTIES, pour la durée du CONTRAT, à utiliser ses marques listées en Annexe 5 et sa dénomination sociale dans le seul cadre de la présentation du partenariat ou du projet qui les lie, en dehors de toute association à un produit ou un service de la SOCIETE. Chaque PARTIE pourra suspendre à tout moment cette autorisation.

Tout autre usage, notamment commercial, de l'ensemble des marques et signes distinctifs de l'une des PARTIES ou identifiant ses laboratoires, n'est pas autorisé.

Les PARTIES ne disposent d'aucun droit pour autoriser un tiers et notamment des distributeurs, à utiliser les marques et nom du CNRS, de ses laboratoires ou de ses chercheurs.

Toute mention des noms des chercheurs employés par les ETABLISSEMENTS doit être préalablement autorisée par l'intéressé et respecter les principes et obligations définis à la présente clause.

Les règles exposées ci-dessus sont également applicables au nom et au logo de l'Université de Poitiers et de la SOCIETE.

ARTICLE 13 - CAHIER DE LABORATOIRE

Le cahier de laboratoire utilisé par le personnel du LABORATOIRE pendant l'ETUDE est la propriété des ETABLISSEMENTS. La SOCIETE peut consulter la partie du cahier du LABORATOIRE relative à l'ETUDE.

ARTICLE 14 - DUREE

Le CONTRAT prend effet le 30 novembre 2021 et est conclu pour une durée de 36 mois.

ARTICLE 15 – RESILIATION DU CONTRAT

Au cas où l'Association Nationale pour la Recherche Technologique (ANRT) suspendrait la subvention CIFRE en raison par exemple d'une interruption notable des travaux, la SOCIÉTÉ s'engage à le faire savoir sans délai aux ETABLISSEMENTS.

Les PARTIES s'engagent à aménager les conséquences de cette suspension par voie d'avenant si celle-ci est susceptible d'entraîner des conséquences préjudiciables sur la réalisation de l'ETUDE. En cas de résolution de la subvention CIFRE par l'ANRT, le CONTRAT est automatiquement résilié à la date d'information du CNRS/ ETABLISSEMENT par la SOCIETE.

Le CONTRAT pourra être résilié de plein droit, par lettre recommandée avec accusé de réception, par l'une ou l'autre des PARTIES du fait de la survenance de l'un des événements suivants :

- en cas de force majeure, selon les modalités prévues à l'article 17 ;
- en cas de (i) violation de l'une quelconque des obligations visées par l'Article 10 « Données personnelles » du CONTRAT, ou par l'article 8 (confidentialité) ou 9 (publications) (ii) violation de la charte SSI du CNRS telle que visée à l'article 18 ;

Le CNRS pourra également résilier de plein droit le CONTRAT selon les modalités prévues par le présent article, en cas de non-paiement des frais et des sommes détaillés à l'article 4 « Modalités de financement » plus de trente (30) jours après leur date d'exigibilité.

Dans tous les cas, la résiliation du CONTRAT produira ses effets à l'égard de toutes les PARTIES.

Les sommes versées et/ou facturées à la date de la résiliation du CONTRAT demeureront acquises et ne pourront donner lieu à répétition.

La résiliation du CONTRAT ne pourra avoir pour effet de modifier les droits des PARTIES sur les RESULTATS et sur les droits de propriété intellectuelle susceptibles de s'appliquer sur les RESULTATS à la date de la résiliation.

Les présentes dispositions ne sont pas exclusives du droit de l'une ou l'autre des PARTIES de faire prononcer en justice la résolution du CONTRAT et/ou de solliciter le versement de tout dommages et intérêts destinés à réparer le préjudice subi.

ARTICLE 16 – INEXECUTION

16.1 Exécution par l'autre Partie ou par un tiers

Les PARTIES conviennent explicitement d'écarter l'application des dispositions de l'article 1222 du Code civil.

Elles ne seront donc pas autorisées à exécuter elles-mêmes ou faire exécuter par un tiers, même sur autorisation préalable du juge, l'obligation mal exécutée ou inexécutée de l'autre PARTIE.

De la même manière, elles ne seront pas autorisées à détruire ce qui a été exécuté par l'autre PARTIE en violation de l'obligation souscrite par celles-ci, y compris à ses frais.

16.2 Suspension d'inexécution

Les PARTIES ne pourront en aucun cas suspendre l'exécution de leur obligation au motif qu'il serait manifeste que l'autre PARTIE n'exécutera pas les siennes dans les conditions prévues par le CONTRAT.

Chaque PARTIE ne sera autorisée à opposer l'exception d'inexécution de ses obligations par l'autre PARTIE qu'à la condition qu'elle soit avérée et suffisamment grave et à la condition de lui avoir fait parvenir une lettre de mise en demeure recommandée avec accusé de réception qui n'aurait pas été suivie d'effet dans les trente (30) jours au moins à compter de sa première présentation.

ARTICLE 17 - FORCE MAJEURE

Aucune des PARTIES ne sera responsable à l'égard d'une autre ni ne sera réputée avoir manqué à ses obligations découlant des stipulations du CONTRAT en cas de défaut ou de retard d'exécution d'une quelconque des stipulations du CONTRAT si le défaut ou le retard est dû à des causes qui échappent au contrôle raisonnable de la PARTIE affectée, ladite cause ne pouvant être raisonnablement prévue lors de la conclusion du CONTRAT, y compris les incendies, inondations, séismes, embargos, guerres, actes de guerre, attentats, insurrections, émeutes, troubles civils, grèves, contre-grèves ou autres perturbations du travail, défauts ou retards de livraison par des fournisseurs exclusifs, catastrophes naturelles ou actions, omissions ou retards d'exécution d'une quelconque autorité publique ou de la PARTIE intéressée, sous réserve, cependant, que la PARTIE ainsi affectée mette en œuvre des efforts raisonnables pour supprimer les causes de défaut d'exécution ou en atténuer les effets et poursuive l'exécution du CONTRAT dès que possible suite à la disparition desdites causes. La PARTIE affectée adressera promptement aux autres PARTIES un avis écrit pour les informer du retard ou du défaut d'exécution qui a lieu en raison de l'événement de force majeure, mentionnant la nature de l'événement, sa durée anticipée et toute mesure prise pour en éviter ou en atténuer les effets. Si l'événement de force majeure (i) dure plus de 60 jours et (ii) a un effet négatif substantiel sur l'exécution des obligations de la PARTIE affectée, les PARTIES non affectées auront le droit de résilier le présent CONTRAT aux termes d'une notification écrite adressée aux autres PARTIES et en respectant un préavis de 30 jours.

Les stipulations précitées s'appliquent également si le défaut ou le retard d'exécution est dû à une épidémie ou à une pandémie et ce même si ladite épidémie ou pandémie pouvait être raisonnablement prévue lors de la conclusion du CONTRAT.

En cas de résiliation, les sommes dues par la SOCIETE au CNRS en contribution à l'ETUDE seront calculées et versées *pro rata temporis* compte tenu de la durée séparant la survenance de l'évènement du terme normal du CONTRAT.

ARTICLE 18 – PRESENCE DU PERSONNEL D'UNE PARTIE DANS LES LOCAUX DE L'AUTRE

La présence de personnel de l'une des PARTIES dans les locaux d'une autre PARTIE, pour les besoins d'exécution du CONTRAT, obéira aux conditions suivantes :

- Ledit personnel devra respecter le règlement intérieur, l'ensemble des règles générales ou particulières d'hygiène et de sécurité en vigueur sur leur lieu d'accueil, la charte de la sécurité des systèmes d'information du CNRS DEC133249DAJ publiée au BO (avril 2014) et de l'autre tutelle de l'Unité (Université de Poitiers), ainsi que toutes les instructions qui lui seront communiquées par la PARTIE accueillante ;
- En tout état de cause, le personnel accueilli demeurera sous l'autorité hiérarchique et disciplinaire de son employeur qui reste également responsable en matière d'assurance et de couverture sociale.

ARTICLE 19 - LOI APPLICABLE

Le CONTRAT est soumis au droit interne français.

ARTICLE 20 – RESOLUTION DES LITIGES

Les PARTIES s'efforceront de résoudre à l'amiable tout litige imputable ou lié au CONTRAT par le recours à la médiation ou à la conciliation. A cette fin, les PARTIES devront désigner un médiateur ou organiser les modalités d'une première réunion de conciliation dans un délai de 15 jours à compter de la demande de l'une des PARTIES d'avoir recours à la médiation ou à la conciliation. Si dans un délai de 3 mois, reconductible une fois sur accord des PARTIES, les PARTIES ne sont pas parvenues à résoudre amiablement le litige par voie de médiation ou de conciliation, le litige pourra être porté devant la juridiction compétente. L'introduction d'une procédure juridictionnelle au mépris des stipulations précitées sera sanctionnée par une irrecevabilité.

Le CONTRAT comprend le présent document et ses Annexes :

1. CONNAISSANCES PROPRES des PARTIES.
2. Description de l'ETUDE
3. Annexe financière
4. Noms et marques

Fait à Poitiers en 2 exemplaires originaux

Pour le CNRS

Date :

07 NOV. 2022

Pour le Président-Directeur Général du CNRS
le Responsable
du Service Partenariat et Valorisation
Limousin Poitou-Charentes
Cédric LEBAILLY

Fait à Montpellier en 2 exemplaires originaux

Pour la SOCIETE

Date : 12-10-2022

Handwritten signature

Namrata PATEL, PhD
Directrice générale, ONAOS
Pépinière d'Entreprise Cap Oméga
Rond-Point Benjamin Franklin
34960 Montpellier Cedex 2
SAS au capital de 30 000€
SIREN : 839 849 080 - APE : 6201Z

Annexe 1 : CONNAISSANCES PROPRES des PARTIES

Connaissances propres de la société LMA :

Le laboratoire LMA possède des connaissances dans le domaine des algorithmes stochastiques, statistiques appliquées aux sciences de la vie, analyse de données, apprentissage statistique des réseaux, statistique bayésienne, classification et intelligence artificielle.

- Le laboratoire a développé le logiciel libfbat (bibliothèque dédiée à l'analyse génétique du pedigree. Robuste à la fois aux erreurs de génotypage et aux données manquantes). Un autre logiciel est en cours de finalisation (BQML : Bayesian, quantile maximum likelihood and maximum likelihood estimation of the ex-Gaussian distribution). Un autre logiciel est en cours de finalisation qui a pour but d'utiliser le reinforcement learning pour des jeux de données à dépendances temporelles complexes (non linéarité, grande dimension).

Connaissances propres de la société ONAOS :

La société ONAOS possède des connaissances dans le domaine de santé numérique, algorithmes d'intelligence artificielle appliqués aux données médicales, traitement automatique du langage naturel, moteurs de raisonnement médical, hybridation d'approches symboliques et stochastiques.

La société a développé :

- le logiciel LOINC MAPPER (outil de transcodage de données biologiques selon la nomenclature internationale LOINC) qui est référencé par l'Agence Nationale de Santé ;
- un annotateur médico-linguistique capable d'extraire automatiquement des informations médicales pertinentes à partir de données médicales non structurées ;
- une base de connaissances médicales comprenant non seulement des concepts médicaux issus de plusieurs ressources, mais aussi des recommandations de sociétés savantes et des règles médicales d'association de concepts et d'interprétation de résultats.

Annexe 2 : Description de l'ETUDE

Titre de la thèse : Application des méthodes d'intelligence artificielle à la prédiction de pathologies liées aux données biologiques des patients hospitalisés.

Sujet de thèse : L'objectif de la thèse est de développer des algorithmes d'analyse et de prédiction de pathologies de patients, basés sur l'intelligence artificielle, à partir de données cliniques et biologiques contenues dans le DPI. Ces dernières étant constituées d'observations cliniques et biologiques non structurées sous forme (1) de textes libres et (2) de données numériques, elles ne sont pas directement exploitables pour un traitement informatique. Par conséquent, l'enjeu de la thèse repose sur le « déverrouillage » de ces informations via les techniques d'intelligence artificielle. Plus précisément, des algorithmes de traitement automatique du langage (TAL) seront développés afin d'exploiter les textes libres du DPI et des algorithmes d'apprentissage automatique seront développés afin de permettre une exploration stochastique des données numériques du DPI. L'ensemble des traitements exploitant le DPI seront testés expérimentalement afin de prédire des pathologies sur un périmètre clinique défini.

Contexte général : Depuis près de 15 ans, nous assistons à l'émergence de deux sources de données prometteuses en santé pour permettre un phénotypage clinico-biologique des individus¹, et leur proposer une médecine dite personnalisée : le séquençage du génome humain (avec des améliorations de la technologie de séquençage) et l'adoption généralisée du dossier patient informatisé (DPI). Pour ces deux sources de données, une grande partie de l'information requise pour effectuer une médecine personnalisée est contenue dans des textes écrits non structurés, dits « libres », tels que la littérature biomédicale et les notes clinico-biologiques des DPI. Dans son état actuel, cette information ne permet pas de traitement informatique. Par conséquent, le « déverrouillage » de ces informations via les traitements automatiques du langage naturel (TALN) par une approche classique ou une approche de réseaux de neurones profonds est un domaine prometteur pour la médecine du futur. Ce « déverrouillage » associé avec des modèles prédictifs entraînés sur des larges volumes de données cliniques permettrait de répondre aux besoins grandissants de systèmes d'aide à la décision médicale [4]

Mots-clés : traitement automatique du langage naturel, fouille de texte, phénotypage clinique, anonymisation, annotations, systèmes informatiques de santé, ontologies médicales, relations ontologiques, réseaux de neurones artificiels, apprentissage automatique, apprentissage profond, induction, inférence, systèmes symboliques, bases de connaissance, base d'apprentissage, apprentissage supervisé, IUC (Identifiant Unique de Concept), LOINC, SNOMED CT, systèmes d'aide à la décision médicale

1. Contexte

1. 1. Industriel

Onaos est une société créée par un médecin et une chercheuse en Intelligence Artificielle (IA), à une époque où l'adoption généralisée du dossier patient informatisé (DPI) favorise la production d'un volume grandissant de données de santé informatisées et où le savoir médical croît de manière exponentielle (les experts prédisent un doublement de ce savoir tous les 73 jours dès 2021)².

L'assistance technologique aux médecins rendue possible par la maturité des techniques

¹ Le phénotypage a été défini comme « la reconnaissance algorithmique de toute cohorte de patients au sein d'un DPI, dans un but défini, incluant, la sélection de cohortes cas-témoins pour des études d'association pangénomique, des essais cliniques, des mesures de qualité (quality data model [1] = modèle d'informations qui définit les relations entre les patients et les concepts cliniques dans un format standardisé & normalisé pour permettre la mesure informatique d'indicateurs qualité) et des supports d'aide à la décision médicale » [2], [3]

² Densen P. Challenges and Opportunities Facing Medical Education. Trans Am Clin Climatol Assoc. 2011;122:48-58

d'intelligence artificielle promet en effet de diminuer le risque d'erreurs médicales, qui provoquent chaque année 251 000 décès aux États-Unis [5]. Parmi les erreurs médicales rencontrées, une étude menée en Australie, Canada, Allemagne, Nouvelle Zélande, Royaume Uni et États Unis montre que selon le pays 4 à 19% des patients analysés ont subi une erreur médicale liée à leurs résultats biologiques. Parmi celles-ci, on notera : un résultat biologique incorrect (3 à 10%), mais surtout un retard à la communication et à la prise en charge d'un résultat biologique anormal (5-16%) [6].

Bien que les coûts liés aux analyses de biologie médicale pèsent peu sur les dépenses de santé de ville (environ 4% en France) leur impact sur le coût des soins en aval est considérable, puisqu'elles participent dans 70% des situations aux diagnostics médicaux. On comprend dès lors comment une mauvaise gestion de ces dernières peut avoir de graves conséquences.

Surajoutés aux erreurs médicales qui leur sont liées, on constate également qu'elles sont souvent surutilisées, avec notamment une redondance des analyses classiques de routine parfois inutile [7]. Enfin, l'introduction récente et continue de nouvelles analyses biologiques de plus en plus coûteuses (génétique, biologie moléculaire) vient ajouter un surcoût significatif lié aux examens biologiques pour l'assurance maladie.

Ceci a conduit l'Assurance maladie, via la Haute Autorité de Santé, à mieux définir les conditions d'éligibilité au remboursement des actes les plus fréquents ou coûteux, rendant la prescription d'examen biologiques plus contraignante pour les cliniciens (exemple vitamine D)³.

Face à ce constat de difficultés d'utilisation et d'exploitation des examens biologiques, Onaos souhaite développer un système d'aide à la prescription d'actes de biologie médicale, basé sur l'extraction automatique des informations cliniques et biologiques disponibles dans le DPI au moment de la prescription. Les informations ainsi extraites serviront d'éléments d'entrée à un moteur algorithmique de règles médicales prenant appui sur les référentiels médicaux de bonne pratique. La mise à jour en temps réel de ce moteur algorithmique garantira au médecin une pratique de prescription optimale et toujours actualisée des examens biologiques.

Parallèlement à l'aide à la prescription d'examen biologiques, Onaos souhaite développer un système d'aide à l'interprétation des résultats biologiques. Ce système transformera un résultat biologique numérique anormal en pathologie potentiellement associée à cette anomalie biologique, diminuant ainsi pour le médecin le temps quotidien nécessaire à l'interprétation des résultats biologiques. Ce système permettra également d'améliorer la qualité de soins, puisqu'il a été démontré qu'un tel système améliore l'efficacité de la surveillance biologique des traitements médicamenteux qui le nécessite [8].

2. 2. Scientifique

Notre programme de R&D représente la vision commune de l'équipe fondatrice d'Onaos : la réalisation d'une intelligence artificielle hybride, capable d'allier les derniers modèles d'apprentissage automatique avec la connaissance acquise par l'expertise médicale, dans l'objectif d'assister le médecin dans son diagnostic clinique.

A l'heure actuelle, une part importante des données médicales n'est pas structurée, et celles qui le sont ne sont pas normalisées. La numérisation de cette connaissance est un sujet d'actualité et des actions sont mises en place dans ce sens, telles que la e- prescription prévue pour 2021-2022. Cette dernière soulève toutefois plusieurs enjeux. En effet, il faut assurer l'interopérabilité des données et des résultats médicaux des patients, et le dossier patient informatisé (DPI) doit être décrit avec des codes lisibles par toute instance médicale.

Notre opération de recherche se décline donc en fonction des questions de recherche suivantes :

1. Comment assurer l'interopérabilité des données médicales ? Notamment, entre les données

³ Assurance Maladie (après avis de la Haute Autorité de Santé), 'Dosage de la vitamine D : Règles de prise en charge par l'assurance-maladie.pdf', Sep. 2014.

cliniques et les données biologiques ;

2. Une fois l'interopérabilité établi :

- a. Comment définir, concevoir et développer un système d'aide à la décision médicale qui permettra d'aider les médecins à choisir le bon examen biologique afin de confirmer une pathologie suspectée cliniquement ?
- b. Comment définir, concevoir et développer un système d'aide à la décision médicale qui permettra d'aider les médecins à interpréter les résultats biologiques en transformant tout résultat biologique numérique anormal en proposition de pathologies reliées à ces anomalies biologiques ?

En réponse à l'étape 1, nous avons développé un outil qui est en phase de commercialisation aujourd'hui. Grâce aux contributions scientifiques issues de notre collaboration avec le CHU de Montpellier dans l'objectif d'évaluer cet outil, nous pouvons nous focaliser sur le développement d'un système d'aide à la décision médicale basé sur des techniques d'intelligence artificielle. **C'est dans ce cadre que s'inscrit cette thèse.**

2. Objectifs

L'objectif de la thèse est de développer des algorithmes d'analyse et de prédiction de pathologies de patients, basés sur l'intelligence artificielle, à partir de données cliniques et biologiques contenus dans le DPI.

Notre défi réside dans le fait que les données contenues dans le DPI ou la littérature biomédicale sont majoritairement constituées de textes écrits non structurés, dits « textes libres ». Dans leur état actuel, ces données ne sont donc pas compatibles avec un traitement informatique. Par conséquent, le « déverrouillage » de ces informations via les techniques d'intelligence artificielle est primordial dans le développement d'un système d'aide à la décision médicale et constitue le cœur de notre activité. C'est également l'objet de la thèse de définir, à partir de critères médico-statistiques, mais aussi technico-économiques, le périmètre clinique des traitements informatiques envisagés.

Des algorithmes de traitement automatique du langage (TAL) seront développés afin d'exploiter les textes libres contenus dans le DPI. Un modèle prédictif basé sur des algorithmes d'apprentissage automatique ainsi que des analyses stochastiques sera également développé. L'ensemble des traitements exploitant le DPI seront testés expérimentalement sur le périmètre clinique défini.

L'extraction automatique d'informations représente une tâche finalisée du TAL. L'utilisation de méthodes d'apprentissage supervisé est l'approche de référence en TAL, et permet d'obtenir les meilleures performances aussi bien sur des tâches intermédiaires d'analyse linguistique (étiquetage en parties du discours, analyse syntaxique, identification de relations de coréférence, détection d'entités nommées, etc.), que pour des tâches finalisées (traduction automatique, dialogue, extraction d'information, etc.) [9]. La mise en œuvre de ces méthodes repose sur la modélisation, sur un corpus d'apprentissage, de corrélations statistiques entre des descripteurs et une ou plusieurs variables d'intérêt, permettant ensuite le calcul (l'inférence) des valeurs de ces variables lorsqu'elles ne sont pas observées.

Cette stratégie est rendue particulièrement difficile en TAL par (a) le besoin de modéliser des phénomènes à partir d'observations parfois très éparpillées et (b) le caractère structuré des entités linguistiques (mots, phrases, documents), qui implique des dépendances souvent complexes, et potentiellement à longue distance, entre les variables que l'on doit prédire.

Depuis l'arrivée en 2018 de « BERT » [10], la solution révolutionnaire du TAL issue de la recherche en apprentissage automatique, et plus précisément en appliquant des techniques de réseaux de neurones, la tâche d'extraction d'informations a atteint un seuil de performance acceptable pour une utilisation dans un contexte généralisé. Cependant, ces performances sont inadéquates dans des domaines techniques dont le langage reste très spécifique, et des travaux supplémentaires de recherche sont

nécessaires pour adapter la solution au domaine spécifique [11].

Bien que l'état de l'art dans l'extraction d'informations a de bonnes performances dans le contexte général, des difficultés ont été identifiées à adapter ces performances au contexte médical notamment à cause du manque de corpus de données médicales de taille conséquente. Il existe également, une difficulté liée à la traçabilité des erreurs dans le raisonnement de l'intelligence artificielle, ce dernier étant accompli par des processus complexes de calculs statistiques. Pour répondre à ces deux enjeux, nous adoptons une approche à base de règles expertes, définies par des médecins et des linguistes, afin de contourner les difficultés liées au manque de corpus de données et de répondre au besoin de traçabilité des erreurs du raisonnement.

Concernant le modèle prédictif basé sur des algorithmes d'apprentissage automatique ainsi que des analyses stochastiques, nous citons trois directions importantes :

- considérer des méthodes d'apprentissage des réseaux, notamment les modèles à bloc stochastiques [12]-[14],
- considérer des modèles de mélange à coefficient variable, afin de pouvoir suivre le profil des patients au cours du temps ainsi de proposer une méthode de classification pour une prise en charge rapide et efficace [15]
- considérer des méthodes d'intelligence artificielle adaptée à la problématique, notamment basées sur des réseaux de neurones récurrents

La thèse cherchera donc à résoudre les verrous suivants :

3. **Comment déterminer les données médicales pertinentes à partir des notes cliniques saisies en texte libre dans le DPI ?** Les modèles d'analyse de langage basés sur des réseaux neuronaux ont de bonnes performances dans le contexte général mais nécessitent des adaptations sur-mesure dans le domaine médical. La thèse cherchera à résoudre cette difficulté dans le contexte médical.
4. **Comment assurer la bonne interprétation médicale par l'algorithme de traitement du langage naturel ?** L'état de l'art dans l'extraction d'informations est basé sur des processus complexes de calculs scientifiques et la disponibilité de larges volumes de données. La thèse cherchera à contourner les difficultés liées au manque corpus de données et de répondre au besoin de traçabilité des erreurs du raisonnement.
5. **Comment exploiter la stabilité et la pertinence de l'intelligence artificielle dans le contexte médical tout en assurant la transparence du raisonnement ?** Dans le domaine médical, cette approche laisse entrevoir des problèmes à solutionner, notamment ceux liés à la compréhension de la logique que l'IA utilise pour produire des recommandations (boîtes noires). La thèse cherchera à concevoir un système hybride qui combine un raisonnement basé sur des règles expertes avec les modèles prédictifs afin de borner les incertitudes de ces derniers.

3. Organisation des travaux de recherche

3.1. Approche méthodologique

Dans un premier temps, une bibliographie sera effectuée afin de dresser un état de l'art suffisant pour orienter les travaux de recherche. Elle portera notamment sur les approches existantes en TAL pour l'extraction automatique des informations dans le domaine médical, les approches prédictives en intelligence artificielle appliquées aux données biologiques et médicales des patients, et le développement des systèmes intelligents d'aide à la décision médicale.

D'un point de vue méthodologique la thèse adressera deux axes distincts d'analyse et de prédiction de pathologies de patients ; le premier axe sera dédié aux données textuelles brutes contenues dans les notes cliniques ; et le deuxième sera dédié aux données biologiques et les résultats numériques contenues

dans le DPI.

Ces deux axes impliqueront les travaux dirigés par la société ONAOS [16] et ceux dirigés par le directeur de thèse Y. SLAOUI, réalisés au LMA de Poitiers [12]-[15]. Ils donneront suite aux travaux réalisés par le candidat Nassar DAKKOUNE dans le cadre d'une alternance au sein d'ONAOS (année universitaire 2020-21) et en collaboration avec Y. SLAOUI. Ces travaux ont été menés sur une base de données patient fonctionnelle, structurée et normalisée, de plus de 40 000 patients. Cette base contient leurs données biologiques, leurs diagnostics d'entrée et de sortie d'hospitalisation, les traitements administrés ainsi que les notes cliniques médicales en texte libre.

En adéquation avec les directions suivies, deux solutions d'IA seront développées et simulées sur (1) les données textuelles contenues dans le DPI et (2) les données cliniques et biologiques, souvent numériques, contenues dans les DPI des patients hospitalisés. Les résultats obtenus permettront de définir un périmètre opérationnel pour ces solutions, afin de concevoir l'architecture nécessaire pour développer un système d'aide à la décision médicale.

En fin de thèse, une expérimentation réelle sera réalisée et validée afin de garantir une industrialisation du procédé.

Afin d'assurer une transition efficace recherche>anticipation>delivery, l'accent sera porté sur les facilités de mise en œuvre de la solution. Par ailleurs, compte tenu de la complexité des environnements réels et des écarts qui peuvent être observés entre simulation et expérience, les solutions seront testées (Proof Of Concept) en collaboration avec les CHU de Montpellier et de Poitiers tout au long de la thèse.

3.2. Environnement de travail

A Onaos, le doctorant sera accueilli dans une composée d'ingénieurs, docteurs et techniciens avec une forte compétence technique et scientifique dans le domaine de biologie, de la médecine et de l'intelligence artificielle. La société accueille également des stagiaires et apprentis.

A l'université de Poitiers, le doctorant sera accueilli au LMA (Laboratoire de Mathématiques et Applications, UMR 7348 du CNRS). Ce laboratoire a pour objectif de promouvoir une recherche appliquée en privilégiant des interactions avec les sciences expérimentales. Il est composé de quatre équipes dont l'équipe de probabilités, statistiques et applications dans laquelle la thèse s'inscrira. Plus spécifiquement, la thèse se déroulera dans l'équipe de probabilités, statistiques et applications dont les thèmes portent sur les algorithmes stochastiques, l'analyse de données, l'apprentissage statistique, la classification et l'intelligence artificielle.

L'encadrement du doctorant sera conjointement assuré par son directeur de thèse académique, le Docteur Yousri SLAOUI, maître de conférences, habilité à diriger des recherches, et ses encadrants industriels, le Dr. Guilhem Mayoral, directeur médical, et la Dr. Namrata PATEL, directrice de recherche et maître de conférences associée.

Dr. Yousri SLAOUI est un expert des algorithmes stochastiques et d'analyse de données, et travaille depuis plusieurs années sur le thème de l'apprentissage statistique, la classification et l'intelligence artificielle.

Les encadrants industriels sont experts chacun de leurs domaines respectifs : la biologie médicale et l'intelligence artificielle et possèdent à ce titre, une reconnaissance de leurs pairs tant en interne qu'en externe.

Au niveau organisationnel, une réunion mensuelle sera organisée par le doctorant pour faire état de l'avancement de ses travaux à ses encadrants industriels et académique. Il préparera à cet effet :

- l'ordre du jour (avec revue d'articles de bibliographie, présentation de résultats, etc.) ;
- le compte-rendu (avec les points d'actions et les inflexions décidées avec les encadrants).

Les résultats de la thèse alimenteront de façon continue rapports internes, publications et colloques.

Le doctorant sera également amené à présenter ses travaux dans la communauté scientifique, au sein de l'école doctorale dont il dépend et plus largement dans toute conférence, chaque fois qu'il sera

sollicité pour cela par ses encadrants.

4. Planning global du déroulement de la thèse (grandes lignes)

	Oct-21	Jan-22	Apr-22	Jul-22	Oct-22	Jan-23	Apr-23	Jul-23	Oct-23	Jan-24	Apr-24	Jul-24
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
Etape 1 - Bibliographie												
Initialisation												
Finalisation (à faire tout au long de la thèse)												
Etape 2 - Description théorique des méthodes proposées												
Description théorique des méthodes proposées adaptée à la problématique												
Etude de sensibilité.												
Livrable présentant les résultats ainsi que les grandeurs/actions à considérer												
Communication scientifique (rédaction d'une publication à minima)												
Etape 3 : Développement d'un package sous forme d'un logiciel												
Première version de l'outil												
Validation expérimentale												
Rédaction de publication dans un congrès ou un journal												
Extension des possibilités de l'outil. (c.f. grandeurs/actions identifiées précédemment)												
Rédaction de publication dans un congrès ou un journal												
Finalisation de l'outil et Proof of Concept												
Etape 4 : Mémoire et soutenance												
Rédaction du mémoire												
Préparation soutenance												

5. Bibliographie

- [1] W. K. Thompson *et al.*, 'An Evaluation of the NQF Quality Data Model for Representing Electronic Health Record Driven Phenotyping Algorithms', *AMIA. Annu. Symp. Proc.*, vol. 2012, pp. 911-920, Nov. 2012.
- [2] J. Pathak *et al.*, 'Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPN consortium', *J. Am. Med. Inform. Assoc. JAMIA*, vol. 20, no. e2, pp. e341-e348, Dec. 2013, doi: 10.1136/amiajnl-2013-001939.
- [3] S. D. R, C. Chris, L. Frank, and E. P. L, 'HTP-NLP: A New NLP System for High Throughput Phenotyping', *Stud. Health Technol. Inform.*, pp. 276-280, 2017, doi: 10.3233/978-1-61499-753-5-276.
- [4] P. I. Dissanayake, T. K. Colicchio, and J. J. Cimino, 'Using clinical reasoning ontologies to make smarter clinical decision support systems: a systematic review and data synthesis', *J. Am. Med. Inform. Assoc.*, vol. 27, no. 1, p. 16, 2020.
- [5] J. G. Anderson and K. Abrahamson, 'Your Health Care May Kill You: Medical Errors.', in *ITCH*, 2017, pp. 13-17.
- [6] C. Schoen *et al.*, 'Taking The Pulse Of Health Care Systems: Experiences Of Patients With Health Problems In Six Countries: Patients' voices can provide policy leaders with a window onto what is happening at the front lines of care.', *Health Aff. (Millwood)*, vol. 24, no. Suppl1, pp. W5-509-W5-525, Jan. 2005, doi: 10.1377/hlthaff.W5.509.
- [7] K. Lewandrowski, 'Integrating Decision Support into a Laboratory Utilization Management Program', *Clin. Lab. Med.*, vol. 39, no. 2, pp. 245-257, Jun. 2019, doi: 10.1016/j.cll.2019.01.004.
- [8] N. S. Whitehead *et al.*, 'The Effect of Laboratory Test-Based Clinical Decision Support Tools on Medication Errors and Adverse Drug Events: A Laboratory Medicine Best Practices Systematic Review', *J. Appl. Lab. Med.*, vol. 3, no. 6, pp. 1035-1048, May 2019, doi: 10.1373/jalm.2018.028019.
- [9] E. Knyazeva, G. Wisniewski, and F. Yvon, 'Les méthodes " apprendre à chercher " en traitement automatique des langues : un état de l'art', *Trait. Autom. Lang. TAL*, vol. 59, no. 1, pp. 39-63, 2018.

- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *ArXiv181004805 Cs*, May 2019, Accessed: May 21, 2021. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [11] Y. Peng, S. Yan, and Z. Lu, 'Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets', *ArXiv190605474 Cs*, Jun. 2019, Accessed: May 21, 2021. [Online]. Available: <http://arxiv.org/abs/1906.05474>
- [12] A. E. Haj, Y. Slaoui, P.-Y. Louis, and Z. Khraibani, 'Estimation in a binomial stochastic blockmodel for a weighted graph by a variational expectation maximization algorithm', *Commun. Stat. - Simul. Comput.*, pp. 1-20, May 2020, doi: 10.1080/03610918.2020.1743858.
- [13] A. Atoui *et al.*, 'Application of the Stochastic Block Model to study the effect of the pollution in the Litani river, Lebanon', *En révision*, 2021.
- [14] A. E. Haj, Y. Slaoui, Z. Khraibani, P.-Y. Louis, and C. Perret, 'Clustering in Attributed Weighted Nodes Network using Stochastic Block Model for Application to EEGData', *En révision*, 2021.
- [15] A. Ounajim, Y. Slaoui, P.-Y. Louis, M. Billot, D. Frasca, and Ph. Rigoard, 'Mixture of varying-coefficient models with random effects processes for intra and inter-subject effect estimation', *En révision*, 2021.
- [16] T. Lemaitre, C. Gosset, M. Lafourcade, N. Patel, and G. Mayoral, 'DEFT 2020 - Extraction d'information fine dans les données cliniques : terminologies spécialisées et graphes de connaissance (Fine-grained Information Extraction in Clinical Data : Dedicated Terminologies and Knowledge Graphs)', in *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Atelier Défi Fouille de Textes, Nancy, France, June 8-19, 2020*, 2020, pp. 55-65. [Online]. Available: <https://www.aclweb.org/anthology/2020.jeptalnrecital-deft.6/>

Annexe 3 : Annexe Financière

POSTES DE DEPENSES	Coût € HT
Coût de personnel permanent - Yousri Slaoui (MCU, 10%)	41 800€ HT
Coût de personnel non permanent (CDD et Stagiaires)	6 400€ HT
Environnement du personnel non permanent	32 000€ HT
Frais de fonctionnement	10 000€ HT
Achat de matériel spécifique	0 € HT
Amortissement de matériel	0 € HT
Frais de mission	10 000€ HT
Sous-traitance	0 € HT
Autres dépenses	0 € HT
Frais de gestion	3 600€ HT
COÛT TOTAL DE L'ETUDE	113 800€ HT
Participation financière du partenaire aux coûts du Laboratoire	30 000€ HT
Reste à payer pour le Laboratoire	83 800€ HT

ANNEXE 4 - NOMS ET MARQUES En vue de l'application de l'article 12 du CONTRAT

Liste des noms, marques signes distinctifs appartenant au CNRS

- Marque verbale française n° 3038307 et européenne n°1735430 « Centre National de la Recherche Scientifique » :
- Marque verbale française n° 3961098 et européenne n° 7244981 « CNRS »
- Marque semi-figurative française n° 4504410 et européenne n° 17994053
- Nom du laboratoire (et logo éventuel)



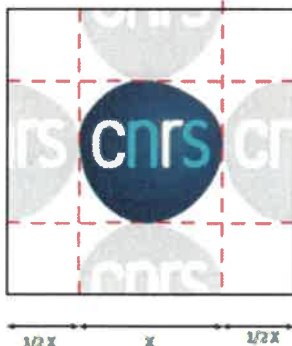
Charte graphique du CNRS

Le logotype CNRS s'utilise seul sans association de typographie. La zone de protection du logotype correspond à une zone vierge autour du logo au moins égal à la moitié de la largeur du logo (ou de sa hauteur).

Le logotype CNRS existe en version quadri et en bleu CNRS (pour des utilisations sur fond clair) et en blanc (pour des utilisations sur fond foncé). Il est strictement prohibé d'utiliser d'autres couleurs.

La taille minimale d'utilisation du logotype est de 12 millimètres.

Zone de protection du logotype



Version quadri



Version bleu
CNRS



Version blanc



Bleu CNRS

C : 100
M : 00
J : 30
N : 00

B : 00
V : 40
B : 15

POURQUOI ?
2000 C

Liste des noms, marques signes distinctifs appartenant à la SOCIETE

Marque verbale française n° 4426699 « Onaos »

Charte graphique d'Onaos

Version classique



Version blanc



Code couleur HEXA

- Bleu : **#4DB6EB**
- Blanc : **#FFFFFF**

Code couleur CMJN

- | | |
|---------------|--------------|
| • Bleu : | • Blanc : |
| C : 67 | C : 0 |
| M : 23 | M : 0 |
| J : 0 | J : 0 |
| N : 8 | N : 0 |